Feature Extraction through Binary Pattern of Phase Congruency for Facial Expression Recognition

Seyedehsamaneh Shojaeilangari
Electrical & Electronics Engg,
Nanyang Tech. Univ.
Singapore

Wei-Yun Yau Jun Li
Institute for Infocomm Research,
A*star,
Singapore

Eam-Khwang Teoh
Electrical & Electronics Engg,
Nanyang Tech. Univ.
Singapore

***Abstract* — Although facial expression plays an important role in human interaction, automated facial expression analysis is still a challenging task. This paper presents a novel facial descriptor based on Phase Congruency (PC) and Local Binary Pattern (LBP) for facial expression recognition. The proposed descriptor, named Binary Pattern of Phase Congruency (BPPC), is an oriented and multi-scale local descriptor that is able to encode various patterns of face images. It is constructed by applying LBP on the oriented PC images. We evaluated the proposed method using the Cohn-Kanade (CK^+) database. In our experiment, we achieved an overall detection rate of 93.83% for the six basic emotions. This shows the effectiveness of the proposed method.**

***Keywords*—facial expression, emotion recognition, phase congruency, local binary pattern**

I. INTRODUCTION

Facial expression is an important cue to convey the human emotional states. This motivates researches to develop systems that can automatically detect and recognize human facial expression. To achieve this, deriving a proper facial pattern representation or feature extraction is crucial.

Local energy model and phase-based approaches have been successfully used to detect various image patterns such as step edges, corners, valleys, and lines. The phase-based model for feature extraction proposes that the useful features of a signal are observed at locations with meaningful Fourier components. This is validated by the human visual system where we perceive image features at points where the phase values of its Fourier components are maximally in congruence [1].

Phase Congruency (PC) based feature detection approaches have some advantages over the gradient based approaches. Unlike the gradient-based approaches such as Sobel operator or Canny edge detector which look for sharp changes of image intensity, PC is a dimensionless quantity. This makes it robust to illumination and contrast variation.

PC based feature detectors have been reportedly used in the field of medical image analysis [2-4]. However, there is only a limited number of reported research works on extracting the low-level facial image features using phase and energy information.

Gundimada and Asari[5] used PC features for face representation and proposed a novel feature selection on PC images for recognition of faces with extreme variations such as partial occlusion, non-uniform occlusion, and varying expressions.

Sarfraz and Hellwich[6] proposed a robust pose estimation procedure based on local energy model for face recognition task. Their novel discriminative feature descriptor is claimed to be insensitive to illumination variation and some other common variations in facial appearance such as skin colour.

The work of Buciu[7] is the only found study on phase congruency for facial expression recognition. The author analyzed the data in frequency domain to measure the similarity of sample phases used to extract reliable features. His experimental results show that their proposed approach is superior to some popular methods used for facial expression recognition.

In this paper, we proposed a novel facial feature descriptor, Binary Pattern of Phase Congruency (BPPC), by combining PC and LBP features in order to capitalize on the advantages of both methods where LBP extracts the textural pattern of the image while PC extracts the discontinuities in the image such as edges, corners etc. Thus the proposed BPPC is an oriented and multi-scale local descriptor that is able to encode various patterns of face images. It is constructed by applying the LBP feature extractor on the oriented PC images so that the patterns of the PC images can be captured. We also used the local phase information of the images to estimate the extracted feature types and then define different LBP codes to describe each feature type.

The novelty of our proposed descriptor are: (1) an oriented and multi-scale local descriptor that is able to encode various patterns of face images (2) use of local phase information to estimate the type of extracted features (3) utilization of different LBP codes for different feature types to represent all facial patterns.

This paper is organized as follows: Section II reviews the background information of PC. Then the proposed feature extraction procedure is explained in detail in section III. Experimental results are discussed in section IV and finally, section V concludes the paper.

II. BACKGROUND INFORMATION OF PHASE CONGRUENCY

For PC calculation, the spatially localized frequency information of an image is required. Wavelet transform is used to achieve this task[4]. To preserve the amplitude and phase information of a signal at a particular frequency, two linear phase quadrature filters are used. In this paper, we used the logarithmic Gabor wavelets as suggested in [4].

Let I denotes the signal and M_n^e and M_n^o represent the even-symmetric and odd-symmetric wavelets at scale *n* respectively. Then the response of each quadrature pair of filters forms a response vector given by:

$$[e_n(x), o_n(x)] = [I(x) * M_n^e, I(x) * M_n^o]. \tag{1}$$

The amplitude and phase information of the transform at a given wavelet scale is formulated as:

$$A_n(x) = \sqrt{e_n(x)^2 + o_n(x)^2}. \tag{2}$$

$$\varphi_n(x) = atan2(e_n(x), o_n(x)). \tag{3}$$

Then the PC is expressed as local energy normalized by the sum of Fourier amplitude components as:

$$PC(x) = \frac{E(x)}{\sum_n A_n(x)}. \tag{4}$$

where $E(x)$ is the local energy defined as:

$$E(x) = \sqrt{F(x)^2 + H(x)^2}. \tag{5}$$

$$F(x) = \textstyle\sum_n e_n(x). \tag{6}$$

$$H(x) = \textstyle\sum_n o_n(x). \tag{7}$$

Taking noise into consideration, the expression of PC can be modified as follows:

$$PC(x) = \frac{\lfloor E(x) - T \rfloor}{\sum_n A_n(x) + \varepsilon}. \tag{8}$$

where the symbol $\lfloor \quad \rfloor$ denotes that the enclosed quantity is not allowed to be negative. T is a noise threshold which is determined from the statistics of the filter response to the data calculated as described in [4] and ε is a small offset to avoid division by zero.

To detect the image features at various orientations, a bank of oriented filters can be designed. However, the multi-orientation representation for the PC calculation may cause high dimensionality which increases the computational time.

III. Methodology

This paper aims to find a feature descriptor that is able to describe various image features at different scales and orientations. Our proposed descriptor is constructed by applying the LBP based structure on the oriented PC images. Combining both PC and LBP methods would allow us to exploit the advantages of both approaches together. Since the PC information can enhance the facial features, especially the region with high energy crucial for facial expression recognition such as edges and lines, it is used as the first step. Then the LBP operator is used as the second step to extract textural facial features where it computes the relative difference in the local structural information and organizing the information using histogram of sub-regions.

To calculate the BPPC for each image pixel, the intensity value required for calculating the traditional LBP is replaced with the PC value at the corresponding pixel. Since the PC image describes different feature types, the LBP code describing each feature type is designed to be different. Additionally, the PC calculation was done across different orientations to describe the image features at various orientations.

The overall process of the proposed BPPC method of feature extraction is shown in Fig. 1. The first step is finding the PC of the image at different orientations. We considered 6 orientations by dividing the frequency plane uniformly. Then, for each PC image, the local phase is estimated to determine the feature type since the local phase denotes a measure of the image feature types. For example, the phase values of ($\pm\pi/2$) indicate the ridge and valley features while (0 or π) determines the step edges. Therefore, after calculating the local phase of each image, we quantized the phase angles into equally sized bins, with each bin representing a particular feature type. The next step is to segment each PC image into spatial patches named "blocks" and finally, the LBP histogram is calculated for each patch. Since different feature types are coded differently, the LBP operator is formed as follow:

$$LBP_{P,R}(x_c, y_c) = Tr(x_c, y_c) + \textstyle\sum_{p=0}^{P-1} S(pc_c, pc_p)\, 2^p \tag{9}$$

where pc_c corresponds to the PC value of the center pixel of a local neighborhood and pc_p($p = 0, \ldots, P-1$) corresponds to the PC value of P equally spaced pixels on a circle of radius R that forms a circularly symmetric set of neighbors. $S(pc_c, pc_p)$ is a sign function and $Tr(x_c, y_c)$ is a threshold function defined for each feature type. For example, if we consider 3 type of features; the function Tr for $LBP_{8,1}$ is defined as follow:

$$Tr(x, y) = \begin{cases} 0 & feature\ type\ I \\ 256 & feature\ type\ II \\ 512 & feature\ type\ III \end{cases}. \tag{10}$$

The BPPC algorithm is summarized in Fig. 2.

The outcomes from different steps of the proposed algorithm are shown in Fig. 3 comprising the original image, local phase, PC, and LBP images respectively. The PC image illustrated in this figure is the combination of 6 oriented PC images. Fig. 3d shows the uniform LBP for 3 feature types. In other word, the phase angles are quantized into 3 equally sized bins which represent ridge, valley, and step edges. Therefore the function Tr is changed to 0, 59, and 118 for feature type I, II, and III respectively.

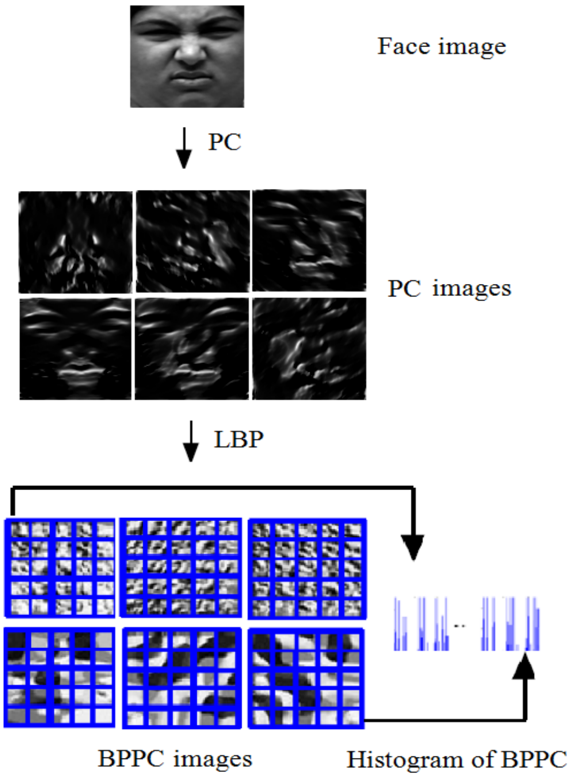

Figure 1. Overview of the proposed method for BPPC feature extraction

BPPC algorithm:

```
Initialize OriNo    %Total No. of orientations
Initialize N        %Total No. of features
Initialize BLKNo %Total No. of blocks
for each orientation (i=1:OriNo)
   Calculate the local phase
   Quantize the phase angle into N equal-sized bins
   Calculate the Phase Congruency (PC)
   Divide the PC image into blocks
for each block (j=1:BLKNo)
for each feature type (k=1:N)
         Calculate LBP histogram for PC block
end
end
   Concatenate the block histograms
end
Concatenate the oriented PC histograms
```

Figure 2. BPPC algorithm

Figure 3. Illustration of the output from different BPPC steps: (a) original sample image, (b) local phase image, (c) PC image, (d) uniform LBP image.

IV. Results and Discussion

We used the extended Cohn-Kanade dataset (CK^+) [8] in this work to evaluate the suitability of our proposed descriptor for facial expression recognition. The database includes 593 video sequences recorded from 123 university students ranging from 18 to 30 years old. The subjects were asked to express a series of 23 facial displays including single or combined action units. Six of the displays were based on descriptions of prototype basic emotions (joy, surprise, anger, fear, disgust, and sadness). Image sequences from neutral to peak emotion display were digitized into 640 by 480 pixels with 8-bit precision for grayscale values. For our work, 309 images of the most expressive faces (last frame of each sequence) which have been labeled as one of six basic emotions are used. Fig. 4 shows the sample images of the basic emotions.

In view of the limited samples, Leave-One subject-Out (LOO) cross validation is used in our experiments. For the database with *N* subjects, *N* experiments were performed. For each experiment, all samples of *N-1* subjects are used for training and the remaining samples for testing. The true detection rate is the average classification accuracy on the test samples. Such an approach also ensures that the results obtained are subject independent since there is no information of the same subject available in both the training and test samples.

We selected the Support Vector Machine (SVM) with polynomial kernel function as the classifier to categorize the samples into six basic emotions. Since SVM is a binary classifier, one-against-all technique was applied for our multi-class problems [9]. Based on this technique, 6 binary SVM classifiers were constructed to categorize each emotion against all the others. The final recognition was obtained using majority voting.

We also performed preprocessing on the original image to prepare the data for feature extraction. The face images were aligned such that the distance between the two eyes in all the images is a constant, followed by rotating the image to align the two eyes' coordinates. Finally, the faces were normalized by cropping them to a rectangle of size 100×100 pixels.

We evaluated the effect of different block sizes and overlapping ratios for the square grid segmentation as shown in Table I. The best results were obtained for blocksize of 4x4 with an overlap ratio of 20% of the original non-overlapping block size.

Our next experiment was performed to evaluate the effect of the number of wavelet scales and their values on classification performance. As Table II shows, the result of 3 scales of the log-Gabor filter with wavelengths of [3 9 27] is superior to the other settings.

The effect of different LBP coding scheme is represented in Table III. We investigated the effect of three types of LBP operator - basic LBP (row 1), uniform LBP (row 2), and rotation invariant uniform LBP (row 3) on the detection rate and the dimension of feature length. The results show that the detection rate of uniform LBP is the best. The result of rotation invariant uniform LBP is also superior if the trade off between memory size and classification accuracy is important since it has the shortest feature length.

To compare the extracted features for different emotion expressions, the BPPC images of all six expressions are depicted in Fig. 5.

We compared our proposed method to the other reported approaches in the literature on the same database in Table IV. Brief information of each of the methods including the number of subjects, number of sequences, evaluation measurement, and detection rate is included. Although our experiment was performed on the largest number of subjects, yet the result is comparable to the best approach reported.

Figure 4. Example of basic emotions from CK+ database. Each image is the last frame of a video clip that shows the most expressive face.

TABLE I. EFFECT OF NUMBER OF BLOCKS AND OVERLAP RATIO OF THE BLOCKS ON FEATURE LENGTH AND DETECTION PERFORMANCE

No. blocks	Overlap ratio (%)	No. features	Detection rate (%)
4 (2×2)	20	4224	84.76
4 (2×2)	50	4224	83.88
16 (4×4)	20	16896	93.83
16 (4×4)	50	16896	88.52
25 (5×5)	20	26400	91.50
25 (5×5)	50	26400	87.25

TABLE II. EFFECT OF DIFFERENT LOG-GABOR FILTER SCALES ON FEATURE LENGTH AND DETECTION PERFORMANCE

No. scales	Filter wavelength	Detection rate (%)
1	[3]	36.11
2	[3 9]	86.95
3	[3 9 27]	93.83
4	[3 9 27 81]	85.6

TABLE III. EFFECT OF DIFFERENT LBP OPERATOR ON FEATURE LENGTH AND DETECTION PERFORMANCE

LBP Type	No. features	Detection rate (%)
$LBP_{8,1}$	24480	89.1
$LBP_{8,1}^{u}$	16896	93.83
$LBP_{8,1}^{riu}$	2784	91.46

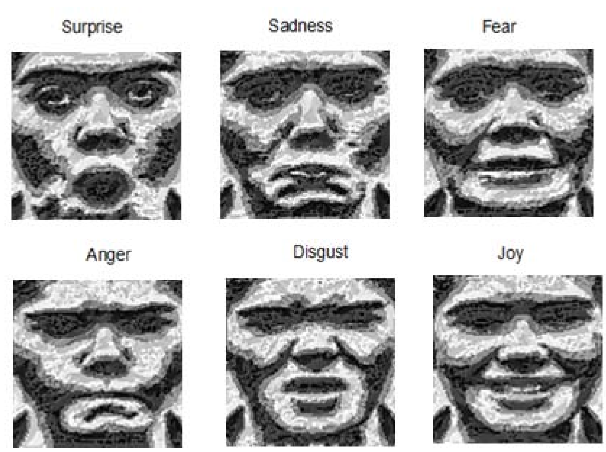

Figure 5. Example of BPPC images for all six emotions.

TABLE IV. COMPARISON OF THE PROPOSED METHOD TO OTHER APPROACHES ON CK^{+} DATABASE

Ref.	# Subject	# Sequence	Evaluation	Recognition Rate (%)
[10]	96	320	10-fold	88.4
[11]	90	313	10-fold	86.9
[12]	90	313	LOO	93.8
[13]	97	-	5-fold	90.9
[14]	90	284	-	93.66
[15]	97	375	-	93.8
[16]	97	375	10-fold	96.26
[17]	115	321	LOO	92.97
Proposed method	118	309	LOO	93.83

V. CONCLUSION

In this paper, we proposed a novel facial feature descriptor for emotion recognition using the combination of PC and LBP features to combine the advantages of both methods together. The oriented and multi-scale local descriptor is able to encode various patterns of facial expression images. Our proposed descriptor was formed by applying the LBP based structure on the oriented PC images to better encode the various patterns of PC images. We also used the local phase information of the images to estimate the extracted feature types and then defined different LBP codes describing each of the feature type. Our experimental results showed that the proposed PC-based model is a useful approach for facial feature representation that may improve the reliability of face emotion recognition system in natural setting. Future works include extending the feature extraction to a video sequence by considering the temporal domain and optimization of the feature extraction process.

REFERENCES

[1] Szilagyi, T. and M. Brady. *Feature extraction from cancer images using local phase congruency: A reliable source of image descriptors*. in *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09*. IEEE International Symposium on. 2009.

[2] Mulet-Parada, M. and J.A. Noble, *2D+T acoustic boundary detection in echocardiography*. Medical Image Analysis, 2000. **4**(1): p. 21-30.

[3] Mellor, M. and M. Brady, *Phase mutual information as a similarity measure for registration*. Medical Image Analysis, 2005. **9**(4): p. 330-343.

[4] Kovesi, P., *Image Features from Phase Congruency*. Videre: Journal of Computer Vision Research, 1999. **1**: p. 1-26.

[5] Gundimada, S. and V.K. Asari, *A Novel Neighborhood Defined Feature Selection on Phase Congruency Images for Recognition of Faces with Extreme Variations*. International Journal of Information Technology, 2006. **3**.

[6] Sarfraz, M.S. and O. Hellwich, *Head pose estimation in phase recognition across pose scenarioss*. International Conference on Computer Vision Theory and Applications VISAPP, 2008. **1**: p. 235–242.

[7] Buciu, I., *Feature extraction through cross-phase congruency for facial expression analysis*. International Journal of Pattern Recognition, 2009. **23**: p. 617-635.

[8] Lucey, P., et al. *The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression*. in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. 2010.

[9] Übeyli, E.D., *Multiclass support vector machines for diagnosis of erythemato-squamous diseases*. Expert Systems with Applications, 2008. **35**(4): p. 1733-1740.

[10] Caifeng, S., G. Shaogang, and P.W. McOwan. *Robust facial expression recognition using local binary patterns*. in Image Processing, 2005. ICIP 2005. IEEE International Conference on. 2005.

[11] Bartlett, M.S., et al. *Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction*. in Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03. Conference on. 2003.

[12] Littlewort, G., et al. *Dynamics of Facial Expression Extracted Automatically from Video*. in Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on. 2004.

[13] Yeasin, M., B. Bullot, and R. Sharma. *From facial expression to level of interest: a spatio-temporal approach*. in Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. 2004.

[14] Asteriadis, S., et al., *Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment*. Multimedia Tools and Applications, 2009. **41**(3): p. 469-493.

[15] Ying-li, T. *Evaluation of Face Resolution for Expression Analysis*. in Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on. 2004.

[16] Guoying, Z. and M. Pietikainen, *Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2007. **29**(6): p. 915-928.

[17] Shojaeilangari, S., Y.W. Yun, and T.E. Khwang. *Person independent facial expression analysis using Gabor features and Genetic Algorithm*. in Information, Communications and Signal Processing (ICICS) 2011 8th International Conference on. 2011.